Zero-Knowledge Against Quantum Attacks

John Watrous
Department of Computer Science
University of Calgary
Calgary, Alberta, Canada
jwatrous@cpsc.ucalgary.ca

ABSTRACT

This paper proves that several interactive proof systems are zero-knowledge against general quantum attacks. This includes the well-known Goldreich-Micali-Wigderson classical zero-knowledge protocols for Graph Isomorphism and Graph 3-Coloring (assuming the existence of quantum computationally concealing commitment schemes in the second case). Also included is a quantum interactive protocol for a complete problem for the complexity class of problems having "honest verifier" quantum statistical zero-knowledge proofs, which therefore establishes that honest verifier and general quantum statistical zero-knowledge are equal: $\mathrm{QSZK} = \mathrm{QSZK}_{\mathrm{HV}}$. Previously no non-trivial proof systems were known to be zero-knowledge against quantum attacks, except in restricted settings such as the honest-verifier and common reference string models. This paper therefore establishes for the first time that true zero-knowledge is indeed possible in the presence of quantum information and computation.

Categories and Subject Descriptors

F.1 [**Computation by Abstract Devices**]: Complexity Measures and Classes, Modes of Computation

General Terms

Theory

Keywords

Zero-knowledge proof systems, quantum cryptography

1. INTRODUCTION

It is clearly to the benefit of honest users of a given cryptographic system that the system is proved secure against as wide a range of malicious attacks as possible. At the same time it is desirable that honest users of the system are subjected to as few resource requirements as possible. The purpose of this paper is to investigate the security of zero-knowledge proof systems against adversaries that use quantum computers to attack these systems. Although quantum interactive proof systems are considered in this paper, the primary focus will be on the case of greatest practical interest, which is the case where honest parties are not required to use quantum computers to implement the proof systems.

The notion of zero-knowledge was first introduced in 1985 by Goldwasser, Micali and Rackoff [14]. Informally speaking, an interactive proof system has the property of being zero-knowledge if verifiers that interact with the honest prover of the system learn nothing from the interaction beyond the validity of the statement being proved. At first consideration this notion may seem to be paradoxical, but indeed several interesting computational problems that are not known to be polynomial-time computable admit zero-knowledge interactive proof systems in the classical setting. Examples include the Graph Isomorphism [11] and Quadratic Residuosity [14] problems, various lattice problems [10], and the Statistical Difference [26] and Entropy Difference [13] problems that concern outputs of boolean circuits with random inputs. (The fact that the last three examples have interactive proof systems that are zero-knowledge relies on a fundamental result of Goldreich, Sahai and Vadhan [12] equating zero-knowledge with "honest verifier" zero-knowledge in some settings.) Under certain intractability assumptions, every language in NP has a zero-knowledge interactive proof system [11].

A related notion is that of an interactive argument, wherein computational restrictions on the prover allow for zero-knowledge protocols having somewhat different characteristics than protocols in the usual interactive proof system setting [3].

There are multiple classical variants of zero-knowledge that differ in the specific way that the notion that the verifier "learns nothing" is formalized. In each variant, it is viewed that a particular verifier learns nothing if there exists a polynomial-time *simulator* whose output is indistinguishable from the output of the verifier when the prover and verifier interact on any positive instance of the problem. The different variants concern the strength of this indistinguishability. In particular, *perfect* and *statistical* zero-knowledge refer to the situation where the simulator's output and the verifier's output are indistinguishable in an information-theoretic sense and *computational* zero-knowledge refers to the weaker restriction that the simulator's output and the verifier's output cannot be distinguished by any computationally efficient procedure.

Within the context of quantum information and computation it is natural to consider the implications of the quantum model to the notion of zero-knowledge. Despite the fact that this has indeed been a topic of investigation for several years, however, relatively little progress has been made. It is straightforward to formulate fairly direct and natural quantum analogues of the definitions of the classical variants of zero-knowledge mentioned above. Known proofs that specific proof systems are zero-knowledge with

STOC'06, May 21–23, 2006, Seattle, Washington, USA.

respect to these classical definitions, on the other hand, do not translate directly to the quantum setting, even when the prover behaves classically. As a result, no nontrivial interactive proof systems had been proved to be zero-knowledge against general quantum attacks previous to this paper. This has left open several possibilities, including the possibility that any "correct" definition of quantum zero-knowledge would necessarily be qualitatively different from the usual classical definitions, as well as the possibility that zero-knowledge is simply impossible in a quantum world.

The main task that needs to be performed to prove that a given proof system is zero-knowledge is the construction of a simulator for every possible deviant polynomial-time verifier. The most typical method involves the simulator treating a given verifier as a *black box*: the simulator randomly generates transcripts, or parts of transcripts, of possible interactions between a prover and verifier, and feeds parts of these transcripts to the given verifier. If the verifier produces a message that is not consistent with the other parts of the transcript that were generated the simulator "rewinds", or backs up and tries again to randomly generate parts of the transcript. By storing intermediate results, and repeating different parts of this process until the given verifier's output is consistent with a randomly generated transcript, the simulation is eventually successful.

The reason why this technique cannot generally be applied directly to quantum verifiers is based on the facts that (i) quantum information cannot be copied, and (ii) measurements are irreversible processes: their effects cannot in general be undone. If a simulator runs a given verifier as a black box and the simulation is unsuccessful, it is not clear how to rewind the process and try again—intermediate states of the system cannot be copied, and running the verifier may have involved an irreversible measurement. More significantly, the determination of whether the simulation was successful will itself represent an irreversible measurement in general. This difficulty was apparently first discussed by van de Graaf [15]. Other methods of constructing simulators for quantum verifiers have also not been successful in the general setting. Further discussions of this issue can be found in [15] and [6].

There are weaker notions of zero-knowledge that are of interest, both in the quantum and classical cases. Of particular interest with respect to previous work on quantum zero-knowledge is the *common reference string* model, wherein it is assumed that an honest third party samples a string from some specified distribution and provides both the prover and verifier with this string at the start of the interaction. Damgård, Fehr, and Salvail [6] proved several interesting results concerning quantum zero-knowledge protocols in this context. Their results are centered on what they call the *no quantum rewinding paradigm*, which partially circumvents the problematic issue concerning simulator constructions discussed above by making use of common reference strings as well as certain unproved quantum complexity-theoretic assumptions. Their results are mostly concerned with interactive arguments, as computational restrictions on the prover are required to establish soundness. Another weaker notion of zero-knowledge is *honest verifier* zero-knowledge, which only requires a simulator that outputs the verifier's view of the interaction between the honest parties V and P. A quantum variant of honest verifier statistical zero-knowledge was considered in [28], wherein it was proved that the resulting complexity class shares many of the basic properties of its classical counterpart [26]. A non-interactive variant of this notion was studied by Kobayashi [21]. The problematic issue regarding simulator constructions does not occur in honest verifier settings.

The present paper resolves, at least to a significant extent, the main difficulties previously associated with quantum analogues of zero-knowledge. This is done by establishing that the most natural quantum analogues of the classical definitions of zero-knowledge indeed can be applied to a large class of proof systems. This includes some well-known classical proof systems as well as quantum proof systems for several problems, in particular the class of all problems admitting quantum proof systems that are statistical zero-knowledge against honest verifiers. It is therefore proved unconditionally that zero-knowledge indeed is possible in the presence of quantum information and computation, and moreover that the notion of quantum zero-knowledge is correctly captured by the most natural and direct quantum analogues of the classical definitions. The main technique used to do this is algorithmic in nature: it is shown how to construct efficient quantum simulators for arbitrary quantum polynomial-time deviant verifiers for several proof systems. These simulators rely on a general *amplification lemma* that establishes simple conditions under which the success probabilities of certain processes with quantum inputs and outputs can be amplified. This process of amplification is similar to one that was previously used to reduce errors in QMA proof systems without increasing witness sizes [22].

2. PRELIMINARIES

This paper assumes the reader is familiar with the notions of interactive proof systems and quantum computation. The results of this paper are most naturally expressed in terms of *promise problems* [7], with which familiarity is also assumed. Notions of zero-knowledge will be discussed briefly, mostly in order to establish notation and to explain the context in which the main results of the paper are relevant. Further information on interactive proof systems and zero-knowledge can be found, for instance, in [8, 9], and standard references for quantum computation and information include [25, 19]. Quantum computational variants of interactive proof systems were studied in [29, 20].

Interactive proof systems

Interactive proof systems will be specified by pairs (V, P) representing honest verifier and honest prover strategies. The soundness property of such an interactive proof system concerns interactions between pairs (V, P') and the zero-knowledge property concerns interactions between pairs (V', P), where P' and V' deviate arbitrarily from P and V, respectively. It may be the case that a given pair of interacting strategies is such that both are classical, both are quantum, or one is classical and the other is quantum. When either or both of the strategies is classical, all communication between them is (naturally) assumed to be classical—only two quantum strategies are permitted to transmit quantum information to one another. It will always be assumed that verifier strategies are represented by polynomial-time (quantum or classical) computations. Depending on the setting of interest, the honest prover strategy P may either be computationally unrestricted or may be represented by a polynomial-time (quantum or classical) computation augmented by specific information about the input string, such as a witness for an NP problem. Deviant prover strategies P' will always be assumed to be computationally unrestricted. (Although the results of this paper are applicable to interactive arguments, none are specific to them, and so for simplicity they are not considered.)

For a given promise problem $A = (A_{\rm yes}, A_{\rm no})$, we say that a pair (V, P) is an interactive proof system for A having completeness error ε_c and soundness error ε_s if (i) for every input $x \in A_{\rm yes}$, the interaction between P and V causes V to accept with probability at least $1 - \varepsilon_c$, and (ii) for every input $x \in A_{\rm no}$ and every prover strategy P', the interaction between P' and V causes V to accept with probability at most ε_s. It may be the case that ε_c and ε_s are constant or are functions of the length of the input string x.

When they are functions, it is assumed that they can be computed deterministically in polynomial time. It is generally desired that ε_c and ε_s be exponentially small. As sequential repetition followed by majority vote, or unanimous vote in case $\varepsilon_c = 0$, reduces these errors exponentially quickly, it is usually sufficient that $1 - \varepsilon_c - \varepsilon_s$ is lower-bounded by the reciprocal of a polynomial. (A similar statement holds for parallel repetition, but the zero-knowledge property to be discussed shortly will generally be lost in this case.)

Zero-knowledge against classical verifiers

There are different notions of what it means for a proof system (V, P) for a promise problem A to be zero-knowledge. Let us first discuss the completely classical case, meaning that only classical strategies are considered for the honest verifier V and any deviant verifiers V'. An arbitrary verifier V' takes two strings as input: a string x representing the common input to both the verifier and prover, as well as a string w called an *auxiliary input*, which is not known to the prover and which may influence the verifier's behavior during the interaction. Based on the interaction with P, the verifier V' produces a string as output. Let $n, m : \{0,1\}^* \to \mathbb{N}$ be polynomially-bounded functions representing the length of the auxiliary input string and output string: assuming the common input string is x, the auxiliary input is a string of length $n(x)$ and the output is a string of length $m(x)$. Because there may be randomness used by either or both of the strategies P and V', the verifier's output will in general be random. The random variable representing the verifier's output will be written $(V'(w), P)(x)$. For the honest verifier V, we may view that $n = 0$ and $m = 1$, because there is no auxiliary input and the output is a single bit that indicates whether the verifier accepts or rejects.

By a (classical) simulator we mean a polynomial-time randomized algorithm S that takes strings w and x, with $|w| = n(x)$, as input and produces some output string of length $m(x)$. Such a simulator's output is a random variable denoted $S(w, x)$. Now, for a given promise problem A, we say that a proof system (V, P) for A is zero-knowledge if, for every verifier V' there exists a simulator S such that $(V'(w), P)(x)$ and $S(w, x)$ are indistinguishable for every choice of strings $x \in A_{\text{yes}}$ and $w \in \{0,1\}^{n(x)}$. The specific formalization of the word "indistinguishable" gives rise to different variants of zero-knowledge. *Statistical* zero-knowledge refers to the situation in which $(V(w), P)(x)$ and $S(w, x)$ have negligible statistical difference, and *computational* zero-knowledge refers to the situation in which no boolean circuit with size polynomial in $|x|$ can distinguish $(V'(w), P)(x)$ and $S(w, x)$ with a non-negligible advantage over randomly guessing. (*Perfect* zero-knowledge is slightly stronger than statistical zero-knowledge in that it essentially requires a zero-error simulation: the simulator may report failure with some small probability, but conditioned on the simulator not reporting failure the output $S(w, x)$ of the simulator is distributed identically to $(V'(w), P)(x)$.)

Two points concerning the definitions just discussed should be mentioned. The first point concerns the auxiliary input, which actually was not included in the definitions given in the very first papers on zero-knowledge (but which already appeared in the 1989 journal version of [14]). The inclusion of an auxiliary input in the definition is needed to prove that zero-knowledge proof systems are closed under sequential composition. Perhaps more important is that the inclusion of auxiliary inputs in the definition captures the notion that a given zero-knowledge proof system cannot be used to *increase* knowledge. The second point concerns the order of quantification between V' and S. Specifically, the definition states that a zero-knowledge proof system is one such that for all V' there exists a simulator S that satisfies the requisite properties. There is a good argument to be made for reversing these quantifiers by requiring that for a given proof system (V, P) there should exist a single simulator S that interfaces in some uniform way with any given V' to produce an output that is indistinguishable from that verifier's output. Typical simulator constructions, as well as the ones that will be discussed in this paper in the quantum setting, do indeed satisfy this stronger requirement.

Zero-knowledge against quantum verifiers

Next let us discuss the case where a given deviant verifier strategy V' may be quantum. This includes the possibility that the honest verifier V is classical or quantum, and likewise for P. Similar to the completely classical case, a verifier V' will take, in addition to the input string x, an auxiliary input, and produce some output. The most general situation allowed by quantum information theory is that both the auxiliary input and the output are quantum states. Moreover, it may be the case that the auxiliary input state qubits are entangled with some other qubits that are not accessible to the verifier or simulator, but are available to any procedure that attempts to distinguish between the verifier and simulator outputs. It is intended that this is a strong assumption, but it can easily be argued that no sensible definition would forbid this possibility; one can imagine natural situations in which potential attacks could be based on entangled states in the sense described.

Similar to the classical case, it will be assumed that for every verifier strategy V' there exist polynomially bounded functions n and m that specify the number of auxiliary input qubits and output qubits of V'. The interaction of V' with P on input x is a physical process, and therefore induces some *admissible* mapping Φ_x from $n(x)$ qubits to $m(x)$ qubits. This means that $\Phi_x : \mathrm{L}(\mathcal{W}) \to \mathrm{L}(\mathcal{Z})$ is a completely positive and trace preserving linear map, where $\mathcal{W}$ and $\mathcal{Z}$ are Hilbert spaces corresponding to the $n(x)$ auxiliary input qubits and the $m(x)$ output qubits, and $\mathrm{L}(\mathcal{W})$ and $\mathrm{L}(\mathcal{Z})$ denote the spaces of linear operators (including the density operators) acting on $\mathcal{W}$ and $\mathcal{Z}$, respectively. Likewise, a simulator S given by some polynomial-time quantum computation that takes as input the string x along with $n(x)$ auxiliary input qubits and outputs $m(x)$ qubits will give rise to some admissible mapping $\Psi_x : \mathrm{L}(\mathcal{W}) \to \mathrm{L}(\mathcal{Z})$.

We may now define variants of zero-knowledge based on different notions of indistinguishability of these mappings Φ_x and Ψ_x. The correct quantum analogue of statistical zero-knowledge requires that $\|\Phi_x - \Psi_x\|_\diamond$ is negligible, where $\|\cdot\|_\diamond$ denotes Kitaev's "diamond" norm [18, 19, 2]. Informally this implies that no physical process can distinguish Φ_x and Ψ_x given a single "black-box" access to one of the two mappings, including the possibility that the mapping is applied to just one part of a larger, possibly entangled state. Under the assumption that $\|\Phi_x - \Psi_x\|_\diamond$ is negligible, no polynomial number of black-box accesses to Φ_x or Ψ_x can suffice to distinguish the two with non-negligible probability. Computational zero-knowledge is formulated similarly, except that the distinguishing procedure must be specified by a polynomial-size quantum circuit. A more precise definition of quantum computational zero-knowledge will be postponed until Section 6.

As in the classical case, a sequential composition of quantum zero-knowledge protocols results in a zero-knowledge protocol, due to the fact that the definitions allow for an arbitrary auxiliary input.

3. THE AMPLIFICATION LEMMA

The polynomial-time quantum simulator constructions for the various protocols considered in this paper rely on a single amplification lemma, stated as Lemma 1 below. The purpose of this section is to explain and prove this lemma.

Suppose that a unitary quantum circuit Q acting on $n + k$ qubits

is given. The first n qubits are assumed to initially store a quantum state $|\psi\rangle$ and the remaining k qubits represent ancillary qubits used by the circuit. If the circuit is applied and the first qubit is measured with respect to the standard basis, the remaining $n+k-1$ qubits will be left in one of two possible states that will be denoted $|\phi_0\rangle$ and $|\phi_1\rangle$, respective to the measurement outcome. This situation is illustrated in Figure 1. Because the states $|\phi_0\rangle$ and $|\phi_1\rangle$, as well

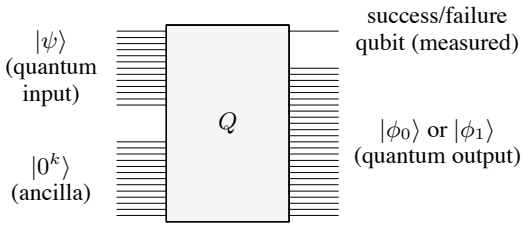

Figure 1: Given circuit Q for amplification lemma

as the probability p to obtain the measurement outcome 0, depend on the choice of $|\psi\rangle$, we will write $p(\psi)$ for p, $|\phi_0(\psi)\rangle$ for $|\phi_0\rangle$, and $|\phi_1(\psi)\rangle$ for $|\phi_1\rangle$ when it is helpful to indicate this dependence explicitly. More concisely, we will write

$$Q\,|\psi\rangle\,|0^k\rangle = \sqrt{p(\psi)}\,|0\rangle\,|\phi_0(\psi)\rangle + \sqrt{1-p(\psi)}\,|1\rangle\,|\phi_1(\psi)\rangle\,,$$

for each possible state $|\psi\rangle$ of the first n qubits.

Let us imagine that it is our goal to construct from Q a procedure that will produce a copy of the state $|\phi_0(\psi)\rangle$ from an arbitrary state $|\psi\rangle$ with as high a success probability as possible (assuming $p(\psi) \neq 0$ so $|\phi_0(\psi)\rangle$ is well defined). This task is of course performed with probability $p(\psi)$ by the circuit Q itself, and without any additional assumptions on Q there may be no way to increase the probability of success in the worst case.

On the other hand, there are some assumptions on Q that do allow for an increase in the probability to obtain $|\phi_0(\psi)\rangle$ given $|\psi\rangle$. The specific assumption that will be required for our amplification lemma is a natural one: that the measurement result gives no information about $|\psi\rangle$. In this case, the probability of successfully obtaining $|\phi_0(\psi)\rangle$ given $|\psi\rangle$ can be made arbitrarily close to 1 by an efficient procedure.

LEMMA 1 (AMPLIFICATION LEMMA). *Let Q be a quantum circuit of the form described above, and assume that the probability $p = p(\psi)$ associated with the measurement outcome 0 is constant over all choices of $|\psi\rangle$ and satisfies $0 < p \leq 1/2$. Then for every $\varepsilon > 0$ there exists a quantum circuit R with $\mathrm{size}(R) = O\left((1/p)\log(1/\varepsilon)\,\mathrm{size}(Q)\right)$ such that for every input state $|\psi\rangle$, R outputs $|\phi_0(\psi)\rangle$ with probability at least $1-\varepsilon$.*

In order to prove Lemma 1, we will make use of a lemma that states a fact first proved in [22], where it was used to analyze an error reduction method for the class QMA.

LEMMA 2. *Let U, Π_0, Π_1, Δ_0, Δ_1 be linear operators acting on some Hilbert space such that U is unitary and Π_0, Π_1, Δ_0, and Δ_1 are orthogonal projections satisfying $\Delta_0 = I - \Delta_1$ and $\Pi_0 = I - \Pi_1$. Suppose further that $|\gamma_0\rangle$ is a unit eigenvector of $\Delta_0 U^* \Pi_0 U \Delta_0$ with corresponding eigenvalue $\lambda \in (0,1)$. Define*

$$|\delta_0\rangle = \frac{\Pi_0 U\,|\gamma_0\rangle}{\sqrt{\lambda}},\ |\delta_1\rangle = \frac{\Pi_1 U\,|\gamma_0\rangle}{\sqrt{1-\lambda}},\ \textit{and}\ |\gamma_1\rangle = \frac{\Delta_1 U^*\,|\delta_0\rangle}{\sqrt{1-\lambda}}.$$

Then $\langle\gamma_0|\gamma_1\rangle = \langle\delta_0|\delta_1\rangle = 0$ and

$$\begin{aligned} U\,|\gamma_0\rangle &= \sqrt{\lambda}\,|\delta_0\rangle + \sqrt{1-\lambda}\,|\delta_1\rangle\,, \\ U\,|\gamma_1\rangle &= \sqrt{1-\lambda}\,|\delta_0\rangle - \sqrt{\lambda}\,|\delta_1\rangle\,. \end{aligned} \tag{1}$$

PROOF. First let us note that because $|\gamma_0\rangle$ is an eigenvector of $\Delta_0 U^* \Pi_0 U \Delta_0$ and the corresponding eigenvalue λ is nonzero, it holds that $\Delta_0\,|\gamma_0\rangle = |\gamma_0\rangle$. By the definitions of $|\gamma_1\rangle$, $|\delta_0\rangle$, and $|\delta_1\rangle$ it also holds that $\Delta_1\,|\gamma_1\rangle = |\gamma_1\rangle$, $\Pi_0\,|\delta_0\rangle = |\delta_0\rangle$, and $\Pi_1\,|\delta_1\rangle = |\delta_1\rangle$. Consequently $\langle\gamma_0|\gamma_1\rangle = \langle\delta_0|\delta_1\rangle = 0$.

The equation $U\,|\gamma_0\rangle = \sqrt{\lambda}\,|\delta_0\rangle + \sqrt{1-\lambda}\,|\delta_1\rangle$ is immediate from the definitions of $|\delta_0\rangle$ and $|\delta_1\rangle$, along with the fact that $\Pi_0 = I - \Pi_1$. Because

$$\frac{\Delta_0 U^*\,|\delta_0\rangle}{\sqrt{\lambda}} = \frac{\Delta_0 U^* \Pi_0 U \Delta_0\,|\gamma_0\rangle}{\lambda} = |\gamma_0\rangle\,,$$

it also holds that $U^*\,|\delta_0\rangle = \sqrt{\lambda}\,|\gamma_0\rangle + \sqrt{1-\lambda}\,|\gamma_1\rangle$, and thus $U\,|\gamma_1\rangle = \sqrt{1-\lambda}\,|\delta_0\rangle - \sqrt{\lambda}\,|\delta_1\rangle$. □

It will be helpful when applying this lemma to note that for U unitary and λ real, the equations (1) are equivalent to

$$\begin{aligned} U^*\,|\delta_0\rangle &= \sqrt{\lambda}\,|\gamma_0\rangle + \sqrt{1-\lambda}\,|\gamma_1\rangle \\ U^*\,|\delta_1\rangle &= \sqrt{1-\lambda}\,|\gamma_0\rangle - \sqrt{\lambda}\,|\gamma_1\rangle\,. \end{aligned}$$

PROOF OF LEMMA 1. For a given circuit Q and error bound ε, let R be a quantum circuit implementing the following algorithm:

Input and initial conditions:

The register W contains an n-qubit quantum input $|\psi\rangle$.
The register X is initialized to the state $|0^k\rangle$.

Main procedure:

Set $t = 0$.

Apply the circuit Q to the pair (W, X) obtaining (B, Y) (where B denotes the first qubit and Y denotes a register containing the remaining $n+k-1$ qubits).

Repeat:

Measure B with respect to the computational basis.

If the outcome of the measurement is 1:

Apply Q^* to (B, Y), obtaining (W, X).

Perform a phase flip in case any of the qubits of X is set to 1. Equivalently, apply the unitary transformation $2\,|0^k\rangle\,\langle 0^k| - I$ to X.

Apply Q to the pair (W, X) obtaining (B, Y).

Set $t = t + 1$.

Until the measurement outcome is 0 or $t = \lceil (1/p)\log(1/\varepsilon)\rceil$.

Output register Y.

In order to analyze this procedure, and in particular to apply Lemma 2, let us define four projections, each acting on $n+k$ qubits:

$$\begin{aligned} \Pi_0 &= |0\rangle\,\langle 0| \otimes I, & \Delta_0 &= I \otimes |0^k\rangle\,\langle 0^k|\,, \\ \Pi_1 &= |1\rangle\,\langle 1| \otimes I, & \Delta_1 &= I - \Delta_0. \end{aligned}$$

The measurement of the qubit B that is performed in the procedure may be viewed as a measurement with respect to the projections $\{\Pi_0, \Pi_1\}$, while the phase flip performed in case the measurement result is 1 may be written $2\Delta_0 - I$. These projections obviously satisfy the conditions $\Delta_0 = I - \Delta_1$ and $\Pi_0 = I - \Pi_1$.

Define $|\gamma_0\rangle = |\psi\rangle\,|0^k\rangle$, where $|\psi\rangle$ is an arbitrary quantum input. It will now be shown that $|\gamma_0\rangle$ is an eigenvector of the operator $\Delta_0 Q^* \Pi_0 Q \Delta_0$, and that the corresponding eigenvalue is p. (This will be so regardless of the choice of $|\psi\rangle$.) Define

$$M = (I \otimes \langle 0^k|)Q^* \Pi_0 Q (I \otimes |0^k\rangle).$$

The operator M may be viewed as a measurement operator on n qubits—the pair $\{M, I - M\}$ describes the POVM-type measurement that is effectively performed on the quantum input $|\psi\rangle$ when the circuit Q is applied and the success/failure qubit is measured. The assumptions of the lemma imply that for *every* choice of the state $|\psi\rangle$ we have $\langle\psi|M|\psi\rangle = \|\Pi_0 Q(|\psi\rangle |0^k\rangle)\|^2 = p$. There is only one possibility for the operator M given this fact: it must be that $M = pI$. This is because M, like any other linear operator, is uniquely determined by the function $|\psi\rangle \mapsto \langle\psi|M|\psi\rangle$ defined on the unit sphere. Consequently,

$$\Delta_0 Q \Pi_0 Q \Delta_0 = (I \otimes |0^k\rangle) M (I \otimes \langle 0^k|) = pI \otimes |0^k\rangle \langle 0^k| .$$

Clearly $|\gamma_0\rangle = |\psi\rangle |0^k\rangle$ is an eigenvector of this operator with corresponding eigenvalue p.

Define

$$|\delta_0\rangle = \frac{1}{\sqrt{p}} \Pi_0 Q |\gamma_0\rangle = |0\rangle |\phi_0\rangle ,$$
$$|\delta_1\rangle = \frac{1}{\sqrt{1-p}} \Pi_1 Q |\gamma_0\rangle = |1\rangle |\phi_1\rangle ,$$
$$|\gamma_1\rangle = \frac{1}{\sqrt{1-p}} \Delta_1 Q^* |\delta_0\rangle .$$

The procedure begins by applying Q to the pair (W, X), which is initially in the state $|\psi\rangle |0^k\rangle = |\delta_0\rangle$. The result is

$$\sqrt{p}\, |0\rangle |\phi_0\rangle + \sqrt{1-p}\, |1\rangle |\phi_1\rangle = \sqrt{p}\, |\delta_0\rangle + \sqrt{1-p}\, |\delta_1\rangle .$$

If the measurement of B results in outcome 0, the state of the pair (B, Y) becomes $|0\rangle |\phi_0\rangle$. The loop is terminated, which results in Y being output in state $|\phi_0\rangle$ as required. If the measurement outcome is 1, the state of (B, Y) becomes $|\delta_1\rangle$.

Consider the effect of the operations performed in case the measurement outcome is 1, assuming the state of (B, Y) is $|\delta_1\rangle$. By Lemma 2, Q^* maps this state to $\sqrt{1-p}\, |\gamma_0\rangle - \sqrt{p}\, |\gamma_1\rangle$, the phase-flip is applied yielding $\sqrt{1-p}\, |\gamma_0\rangle + \sqrt{p}\, |\gamma_1\rangle$, and finally Q is applied resulting in the state $2\sqrt{p(1-p)}\, |\delta_0\rangle + (1-2p)\, |\delta_1\rangle$. Now a measurement of B results outcome 0 and corresponding state $|\delta_0\rangle = |0\rangle |\phi_0\rangle$ with probability $4p(1-p)$ and outcome 1 and corresponding state $|\delta_1\rangle = |1\rangle |\phi_1\rangle$ with probability $(1-2p)^2$. For each subsequent iteration of the loop, which is only performed in case the measurement outcome was 1, the pattern is identical. Consequently, whenever the measurement outcome is 0, the output of the procedure is $|\phi_0\rangle$, and the probability that the measurement outcome is 0 within t iterations is $1 - (1-p)(1-2p)^{2t}$. The probability that $|\phi_0\rangle$ is output by the procedure is therefore greater than $1 - \varepsilon$ if $\lceil (1/p) \log(1/\varepsilon) \rceil$ iterations of the loop are permitted. □

Remark on the connection to Grover's algorithm

The amplification procedure described in the proof of Lemma 1 has some connections with Grover's Algorithm [16] and the more general process known as amplitude amplification [4]. Specifically, if the measurement of B was replaced with a phase-flip $2\Pi_0 - I$, and the loop was terminated after some number of iterations depending on p, then we would essentially be performing amplitude amplification with a quantum state input. This approach can be made to give some reduction in the number of iterations required: essentially replacing p with $\sqrt{p}$. The dependence on ε, on the other hand, does not improve [5].

Although it may be an interesting question to explore the process of amplitude amplification with a quantum state input in other contexts, the author views that it is not helpful in the present case. The main reason is that both the procedure and the analysis generally become more complicated if one wishes to allow ε to be exponentially small, which will be the typical case in the context of zero-knowledge. Because the purpose of a simulator for a zero-knowledge protocols is to establish the security of the protocol, as opposed to being an algorithm that performs a task that is useful in its own right, it seems that it is not worth sacrificing simplicity for performance.

An Amplification Lemma with negligible perturbations

The assumptions of Lemma 1 require that p is independent of $|\psi\rangle$. Here we note that this assumption may be relaxed slightly if one is willing to accept a small perturbation in the output of the procedure.

Suppose that for some circuit Q we have that $p(\psi) \in (0,1)$ for all $|\psi\rangle$, and for some choice of $\delta > 0$ and $q \in (0,1)$ it holds that $|p(\psi) - q| < \delta$ for all $|\psi\rangle$. Then there necessarily exists a unitary operator U that essentially represents an idealized version of Q:

$$U |\psi\rangle |0^k\rangle = \sqrt{q}\, |0\rangle |\phi_0(\psi)\rangle + \sqrt{1-q}\, |1\rangle |\phi_1(\psi)\rangle$$

for all $|\psi\rangle$. Moreover, this unitary operator may be chosen so that $\|Q - U\| < \sqrt{2\delta}$. To prove the existence of such an operator U, one may define a linear operator that acts as required on an orthonormal collection of eigenvectors of $(I \otimes \langle 0^k|) Q^* \Pi_0 Q (I \otimes |0^k\rangle)$, and establish that it is unitary and satisfies the required properties for all choices of $|\psi\rangle$.

Now, if the amplification procedure from the proof of Lemma 1 is applied to the circuit Q, the output of the procedure will have trace distance at most $2\sqrt{2\delta}(2\lceil (1/q)\log(1/\varepsilon)\rceil + 1)$ from what the output would be using the idealized operator U in place of Q. In the situation that δ is exponentially small and $(1/q)\log(1/\varepsilon)$ is polynomially bounded in some input parameter, we have that the output of the procedure for the circuit Q has negligible trace distance from the output of the procedure for the idealized operator U. It will be convenient to use this fact later when quantum statistical and computational zero-knowledge are discussed.

4. GRAPH ISOMORPHISM

The Goldreich-Micali-Wigderson Graph Isomorphism protocol [11] is a simple and well-known example of an interactive proof system that is perfect zero-knowledge against classical polynomial-time verifiers. In this section it is proved that this protocol is zero-knowledge against polynomial-time quantum verifiers as well. The protocol is as follows:

Zero-Knowledge Protocol for Graph Isomorphism

The input is a pair (G_0, G_1) of simple, undirected n-vertex graphs. It is assumed that the prover knows a permutation $\sigma \in S_n$ that satisfies $\sigma(G_1) = G_0$ if G_0 and G_1 are isomorphic.

Prover's step 1: Choose $\pi \in S_n$ uniformly at random and send $H = \pi(G_0)$ to the verifier.

Verifier's step 1: Choose $a \in \{0,1\}$ uniformly at random and send a to the prover. (Implicitly, the verifier is challenging the prover to exhibit an isomorphism between G_a and H.)

Prover's step 2: Set $\tau = \pi\sigma^a$ and send τ to the verifier. (If $\sigma(G_1) = G_0$, then $\tau(G_a) = H$.)

Verifier's step 2: Accept if $\tau(G_a) = H$, reject otherwise.

This proof system has perfect completeness and soundness error 1/2; if $G_0 \cong G_1$, then V will accept with probability 1 when interacting with the honest prover P, while if $G_0 \ncong G_1$ then no prover

P' can convince V to accept with probability greater than 1/2 (essentially because H cannot be isomorphic to both G_0 and G_1 when $G_0 \not\cong G_1$).

For an arbitrary choice of $\sigma \in S_n$ satisfying the required property $\sigma(G_1) = G_0$ for $G_0 \cong G_1$, the proof system (V, P) is perfect zero-knowledge with respect to any classical polynomial-time verifier V'. Sequential repetition followed by a unanimous vote can be used to decrease the soundness error to an exponentially small quantity while preserving the perfect completeness and classical zero-knowledge properties.

We wish to show that this protocol is zero-knowledge with respect to polynomial-time *quantum* verifiers. It will be sufficient to consider a restricted type of verifier as follows:

- In addition to (G_0, G_1), the verifier takes a quantum register W as input, representing the auxiliary quantum input. The verifier will use two additional quantum registers that function as work space: V, which is an arbitrary (polynomial-size) register, and A, which is a single qubit register. The registers V and A are initialized to their all-zero states before the protocol begins.
- In the first message, the prover P sends an n-vertex graph H to the verifier. For each graph H there corresponds a unitary operator V'_H that the verifier applies to the registers $(\mathsf{W}, \mathsf{V}, \mathsf{A})$. After applying the appropriate transformation V'_H, the verifier measures the register A with respect to the standard basis, and sends the resulting bit a to the prover.
- After the prover responds with some permutation $\tau \in S_n$, the verifier outputs the registers $(\mathsf{W}, \mathsf{V}, \mathsf{A})$, along with the classical messages H and τ sent by the prover during the protocol.

An arbitrary verifier can be modeled as a verifier of this restricted form followed by some polynomial-time post-processing of this verifier's output. The same post-processing can be applied to the output of the simulator that will be constructed for the given restricted verifier. Note that a verifier of this form is completely determined by the collection $\{V'_H\}$.

Now let us consider the admissible transformation induced by an interaction of a verifier of the above type with the prover P in the case that $G_0 \cong G_1$. Although the messages sent from the prover to the verifier are classical messages, it will simplify matters to view them as being stored in quantum registers denoted P_1 and P_2, respectively. (Later, when we consider simulations of the interaction, we will need quantum registers to store these messages anyway, and it is helpful to have the registers used in the actual protocol and in the simulation share the same names.) With each register we associate a Hilbert space, and use the same letter in sans serif and calligraphic fonts for matching registers and spaces. For example, $\mathcal{W}$ is the space associated with register W. Let $|0_{\mathcal{V}\otimes\mathcal{A}}\rangle \in \mathcal{V} \otimes \mathcal{A}$ denote the initial all-zero state of the registers (V, A). Let us also write $\mathcal{G}_n$ to denote the set of all simple, undirected graphs having vertex set $\{1, \ldots, n\}$.

For each $H \in \mathcal{G}_n$ and each $a \in \{0, 1\}$, define a linear mapping

$$M_{H,a} = (I_{\mathcal{W}\otimes\mathcal{V}} \otimes \langle a|)\, V'_H (I_{\mathcal{W}} \otimes |0_{\mathcal{V}\otimes\mathcal{A}}\rangle)$$

from $\mathcal{W}$ to $\mathcal{W} \otimes \mathcal{V}$. If the initial state of the register W is a pure state $|\psi\rangle \in \mathcal{W}$, then the state of the registers $(\mathsf{W}, \mathsf{V}, \mathsf{A})$ after the verifier applies V'_H is

$$(M_{H,0}\,|\psi\rangle)\,|0\rangle + (M_{H,1}\,|\psi\rangle)\,|1\rangle\,,$$

and therefore the state of the registers $(\mathsf{W}, \mathsf{V}, \mathsf{A})$ after the verifier applies V'_H and measures A in the standard basis is

$$\sum_{a\in\{0,1\}} M_{H,a}\,|\psi\rangle\,\langle\psi|\,M^*_{H,a} \otimes |a\rangle\,\langle a|\,.$$

The admissible map that results from the interaction is now easily described by incorporating the description of P. It is given by

$$\Phi(X) = \frac{1}{n!} \sum_{\pi\in S_n} \sum_{a\in\{0,1\}} M_{\pi(G_0),a} X M^*_{\pi(G_0),a} \otimes |a\rangle\,\langle a| \otimes |\pi(G_0)\rangle\,\langle\pi(G_0)| \otimes |\pi\sigma^a\rangle\,\langle\pi\sigma^a| \tag{2}$$

for all $X \in \mathrm{L}(\mathcal{W})$.

In order to define a simulator for a given quantum verifier V', it is helpful to consider the classical case. A classical simulation for a classical verifier V' in the above protocol may be obtained as follows. The simulator randomly chooses a permutation π and a bit b, and feeds $\pi(G_b)$ to V'. This verifier chooses a bit a for its message back to the prover. If $a = b$, the simulator can easily complete the simulation, otherwise it "rewinds" and tries a new choice of π and b. With very high probability, the simulator will succeed after no more than a polynomial number of steps, and in the case of success the output of the simulator and the verifier V' will be identically distributed.

Now we consider the quantum case. Our procedure for simulating the verifier described by a collection $\{V'_H : H \in \mathcal{G}_n\}$ will require two registers B and R in addition to W, V, A, P_1, and P_2. The register R may be viewed as a quantum register whose basis states correspond to the possible random choices that a typical classical simulator would use. In the present case this means a random permutation together with a random bit. The register B will represent the simulator's "guess" for the verifier's message. For convenience, let us define $\mathcal{X} = \mathcal{V} \otimes \mathcal{A} \otimes \mathcal{Y} \otimes \mathcal{B} \otimes \mathcal{Z} \otimes \mathcal{R}$, which is the Hilbert space corresponding to all registers aside from W, and let $|0_{\mathcal{X}}\rangle$ denote the all-zero state of these registers.

The procedure will involve a composition of a few operations that we now describe. First, let T be any unitary operator acting on registers $(\mathsf{P}_1, \mathsf{B}, \mathsf{P}_2, \mathsf{R})$ that maps the initial all-zero state of these four registers to the state

$$\frac{1}{\sqrt{2n!}} \sum_{b\in\{0,1\}} \sum_{\pi\in S_n} |\pi(G_b)\rangle\,|b\rangle\,|\pi\rangle\,|\pi, b\rangle\,.$$

If the register R is traced out, the state of registers $(\mathsf{P}_1, \mathsf{B}, \mathsf{P}_2)$ corresponds to a probability distribution over triples $(\pi(G_b), b, \pi)$ for b and π chosen uniformly. In essence, T produces a purification of a uniform distribution of possible *transcripts* of an interaction between a prover and verifier. Next, define a unitary operator V' acting on registers $(\mathsf{W}, \mathsf{V}, \mathsf{A}, \mathsf{P}_1)$ that effectively simulates (unitarily) the verifier V'. Specifically, V' uses P_1 as a control register, and applies V'_H to registers $(\mathsf{W}, \mathsf{V}, \mathsf{A})$ for each possible graph $H \in \mathcal{G}_n$ representing a standard basis state of P_1. More compactly,

$$V' = \sum_{H\in\mathcal{G}_n} V'_H \otimes |H\rangle\,\langle H|\,.$$

The operators T and V' are each tensored with the identity on the remaining spaces when we wish to view them both as operators on $\mathcal{W} \otimes \mathcal{X}$.

Consider the quantum circuit Q acting on all of the above registers obtained by first applying T, then applying V', and finally performing a controlled-NOT operation on the pair (A, B) with A acting as the control. Suppose that Q is applied to $|\psi\rangle\,|0_{\mathcal{X}}\rangle$, and the register B is then measured with respect to the standard basis. The probability that the outcome is 0 is necessarily equal to 1/2, independent of the behavior of the verifier V' and of the auxiliary input $|\psi\rangle$. This follows from similar reasoning to the classical case: there can be no correlation between the verifier's choice of a and the simulator's guess b for a. If we condition on the measurement outcome being 0, and trace out the register R, we obtain precisely

the admissible mapping Φ given in (2) describing the actual interaction between V' and P. In other words, conditioned on the measurement outcome being 0, the circuit Q correctly simulates the interaction between V' and P given auxiliary input $|\psi\rangle$. Given that the measurement outcome is 0 with probability 1/2, which is independent of $|\psi\rangle$, we may apply Lemma 1 in order to obtain a circuit R representing the final simulation procedure.

We note that in the special case $p = 1/2$, the simulation procedure in fact works perfectly after either zero or one iteration of the loop in the amplification procedure. This establishes that the outcome of the simulation procedure is *precisely* $\Phi(|\psi\rangle\langle\psi|)$ in case the initial state of W was $|\psi\rangle$.

Because the set $\{|\psi\rangle\langle\psi| : |\psi\rangle \in \mathcal{W}, \||\psi\rangle\| = 1\}$ spans all of $\mathrm{L}(\mathcal{W})$, and the map induced by the simulation procedure is necessarily admissible (and therefore linear), it holds that this map is precisely Φ. In other words, because admissible maps are uniquely determined by their action on pure states, the map induced by the simulation procedure must be Φ; the simulation procedure implements *exactly* the same admissible map as the actual interaction between V and P.

Each of the operations constituting the circuit Q can be performed by polynomial-size circuits, and therefore the simulator has polynomial size (in the worst case).

5. QUANTUM SZK

The method that was used to prove the security of the Goldreich-Micali-Wigderson Graph Isomorphism protocol against quantum attacks can easily be adapted to some other protocols having a similar form, meaning (i) P sends a message to V, (ii) V flips a fair coin and sends the result to P, and (iii) P responds with a second message. The important aspects of such protocols that may allow the same proof to go through with very little change is that in each case there exists a simulator whose success probability is independent (or nearly independent) of the auxiliary input state of any cheating quantum verifier.

In the quantum setting, protocols of this simple form are universal for honest-verifier quantum statistical zero-knowledge [28], meaning that every problem having a quantum interactive proof system that is statistical zero-knowledge with respect to an honest verifier also has a proof system of the above form. Although such proof systems require the prover to send quantum information to the verifier, and the verifier performs a quantum computation at the end of the protocol, the verifier's single-bit message is classical. (The honest prover can easily enforce this constraint just by measuring the verifier's message before responding to it.) This allows the proof from Section 4 to be easily adapted to this setting as well, provided we use the approximate version of the amplification lemma mentioned in Section 3.

Let $\mathrm{QSZK}_{\mathrm{HV}}$ denote the class of problems having honest-verifier quantum statistical zero-knowledge protocols and QSZK the class of problems that are quantum statistical zero-knowledge with respect to the definitions we have discussed in Section 2. The following theorem then results.

THEOREM 3. $\mathrm{QSZK} = \mathrm{QSZK}_{\mathrm{HV}}$.

Although the statement of this theorem is analogous to the fact $\mathrm{SZK} = \mathrm{SZK}_{\mathrm{HV}}$ of Goldreich, Sahai, and Vadhan [12], we hasten to add that the facts are only really similar on the surface—there is no similarity in the proofs. The quantum case is greatly simplified by the fact that every problem in $\mathrm{QSZK}_{\mathrm{HV}}$ has the very simple type of protocol discussed above.

Because $\mathrm{SZK} \subseteq \mathrm{QSZK}_{\mathrm{HV}}$, all problems in SZK have *quantum* interactive proof systems that are statistical zero-knowledge against quantum verifiers. The question of whether every problem in SZK has a *classical* proof system that is zero-knowledge against quantum attacks is not answered in this paper.

6. QUANTUM COMPUTATIONAL ZERO-KNOWLEDGE AND 3-COLORING

The final proof system that will be discussed is the Goldreich-Micali-Wigderson Graph 3-Coloring proof system [11]. This proof system is computational zero-knowledge against classical verifiers, assuming the existence of unconditionally binding and computationally concealing *commitment schemes* (which follow from the existence of one-way functions [24, 17]). In this section it is argued that this protocol is computational zero-knowledge against quantum verifiers, albeit with somewhat stronger intractability assumptions than are required in the classical case. Specifically, the protocol will require commitment schemes that are unconditionally binding and *quantum* computationally concealing, therefore ruling out schemes based on the computational hardness of factoring, discrete logarithm computations, or any other problem solvable in polynomial time on a quantum computer.

A zero-knowledge proof system for Graph 3-Coloring yields a zero-knowledge proof for any problem in NP, as a protocol for an arbitrary NP problem can begin with both parties computing a reduction to 3-Coloring. The fact that the zero-knowledge property is preserved under such a reduction is discussed in [11], and the quantum and classical settings do not differ in this respect.

We will begin by discussing quantum computational indistinguishability and definitions of both quantum computational zero-knowledge and of quantum computationally concealing commitment schemes. After this, the security of the Goldreich-Micali-Wigderson 3-Coloring proof system against quantum verifiers will be argued. Due to space limitations, this section is not as detailed as a proper discussion of its subject requires, and should therefore be viewed as a preliminary sketch that will be expanded in the final version of this paper.

It will be helpful for the discussions that follow that some conventions and notations regarding quantum circuits are mentioned at this point. We will allow quantum circuits to include two simple, non-unitary gates: *ancillary* gates, which take no input and output a single qubit in state $|0\rangle$, and *trace-out* gates that take one input qubit and give no output, effectively throwing the qubit in the trash. In addition to these two gates, quantum circuits may include Toffoli gates, Hadamard gates, and imaginary-phase-shift gates (which induce the transformation $|0\rangle \mapsto |0\rangle$ and $|1\rangle \mapsto i\,|1\rangle$). By taking these five gates as a basis, we have a universal collection, by which it is meant that an arbitrary admissible quantum operation can be approximated to any desired accuracy by some quantum circuit. Obviously, a quantum circuit may therefore have a different number of input and output qubits; we will say that a circuit is of type (n, m) if it has n input qubits and m output qubits. More generally, an arbitrary admissible map from n qubits to m qubits will be said to be of type (n, m). The *size* of a type (n, m) quantum circuit is defined to be the number of gates in the circuit plus n (to disallow the possibility that a tiny circuit acts on a large number of qubits). When Q is such a circuit, we identify Q with the admissible map from n qubits to m qubits induced by running Q.

Quantum computational indistinguishability

A quantum analogue of computational zero-knowledge requires a formal notion of quantum computational indistinguishability. The definition that follows represents the notion that will be considered in the remainder of this section.

DEFINITION 4. Assume that $S \subseteq \{0,1\}^*$ is an infinite set of strings, $m : \{0,1\}^* \to \mathbb{N}$ is a polynomially bounded function, and ρ_x and ξ_x are mixed states on $m(x)$ qubits for each $x \in S$. The ensembles $\{\rho_x : x \in S\}$ and $\{\xi_x : x \in S\}$ are *polynomially quantum indistinguishable* if, for every choice of

1. polynomials p and q,
2. a polynomially-bounded function $k : \{0,1\}^* \to \mathbb{N}$,
3. a collection $\{\sigma_x : x \in S\}$, where σ_x is a mixed state on $k(x)$ qubits, and
4. a quantum circuit Q of type $(m(x) + k(x), 1)$ and size at most $p(|x|)$,

it holds that

$$|\langle 1|Q(\rho_x \otimes \sigma_x)|1\rangle - \langle 1|Q(\xi_x \otimes \sigma_x)|1\rangle| < \frac{1}{q(|x|)}$$

for all but finitely many $x \in S$.

If $\{\rho_n : n \in \mathbb{N}\}$ and $\{\xi_n : n \in \mathbb{N}\}$ are ensembles indexed by the natural numbers, we identify S with 1^*, interpreting each n with its unary representation. Let us also note that the above definition applies to the situation where $\{\rho_x : x \in S\}$ and $\{\xi_x : x \in S\}$ represent classical probability distributions, which are special cases of mixed states.

Notice that the above definition gives a fairly strong quantum analogue to the typical non-uniform notion of classical polynomial indistinguishability. It is strong because the non-uniformity includes an *arbitrary* quantum state σ_x that may aid some circuit Q in the task of distinguishing ρ_x from ξ_x. The inclusion of the arbitrary state σ_x is important in situations, such as those we will consider in the context of zero-knowledge, where indistinguishability of two ensembles must hold in the presence of auxiliary quantum information.

Next let us extend this definition to admissible mappings. This is done by simply considering ensembles that result from applying the mappings to arbitrary polynomial-size states.

DEFINITION 5. Assume that $S \subseteq \{0,1\}^*$ is an infinite set of strings, $n, m : \{0,1\}^* \to \mathbb{N}$ are polynomially bounded functions, and Φ_x and Ψ_x are admissible mappings of type $(n(x), m(x))$ for each $x \in S$. The ensembles $\{\Phi_x : x \in S\}$ and $\{\Psi_x : x \in S\}$ are *polynomially quantum indistinguishable* if, for every choice of

1. polynomials p and q,
2. a polynomially bounded function $k : \{0,1\}^* \to \mathbb{N}$,
3. a collection of mixed states $\{\sigma_x : x \in S\}$, where σ_x is a state on $n(x) + k(x)$ qubits, and
4. a quantum circuit Q of type $(m(x) + k(x), 1)$ and size at most $p(|x|)$,

it holds that

$$|\langle 1|Q((\Phi_x \otimes I)(\sigma_x))|1\rangle - \langle 1|Q((\Psi_x \otimes I)(\sigma_x))|1\rangle| < \frac{1}{q(|x|)}$$

for all but finitely many $x \in S$.

Note that a slight simplification is incorporated into this definition: the input state σ_x to the admissible mappings may include a part that aids a given circuit Q in distinguishing the outputs.

Now we are prepared to state a definition for quantum computational zero-knowledge. Let (V, P) be a proof system (quantum or classical) for a promise problem $A = (A_{\text{yes}}, A_{\text{no}})$. This proof system will be said to be a *quantum computational zero-knowledge* for A if, for every polynomial-time quantum verifier V' there exists a polynomial-time quantum algorithm $S_{V'}$ that satisfies the following requirements. Assume that on input x, the verifier V' takes $n(x)$ auxiliary input qubits and outputs $m(x)$ qubits, and let Φ_x denote the admissible mapping of type $(n(x), m(x))$ that results from the interaction of V' with P. Then the simulator $S_{V'}$ must also take $n(x)$ qubits as input and output $m(x)$ qubits, thereby implementing a mapping Ψ_x of type $(n(x), m(x))$. Moreover, the ensembles $\{\Phi_x : x \in A_{\text{yes}}\}$ and $\{\Psi_x : x \in A_{\text{yes}}\}$ must be polynomially quantum indistinguishable.

Quantum computationally concealing commitments

Next, we consider commitment schemes that are secure against quantum attacks. It is well-known that there cannot exist unconditionally binding and concealing commitments based on quantum information alone [23], and therefore one must consider commitments for which either or both of the binding and concealing properties is based on a computational assumption. In the interactive proof system setting, where one requires soundness against arbitrary provers, the binding property of the commitments must be unconditional, and therefore the concealing property must be computationally-based.

Naturally, to be secure against quantum attacks, the commitment scheme that is used must in fact be *quantum* computationally concealing. The existence of such schemes has not been proved, and does not follow from the existence of classical computationally concealing commitment schemes. For example, good candidates for classically secure schemes based on the computational difficulty of factoring or computing discrete logarithms become insecure in the quantum setting because of Shor's algorithm [27]. Classical commitments can, however, be based on arbitrary one-way functions [24, 17], and there are candidates for such functions that may be difficult to invert even with efficient quantum algorithms. Functions based on lattice problems, error-correcting codes, and non-abelian group-theoretic problems represent candidates.

DEFINITION 6. Assume that Γ is a finite set with $|\Gamma| \geq 2$. An *unconditionally binding, quantum computationally concealing Γ-commitment scheme* consists of a deterministic polynomial-time computable function f with the following properties.

1. *(Uniform length.)* For some polynomial p we have $|f(a,x)| = p(|x|)$ for every $a \in \Gamma$ and $x \in \{0,1\}^*$. (This requirement is not really essential, and is only made for convenience.)
2. *(Binding property.)* For every choice of $a \neq b \in \Gamma$ and $x, y \in \{0,1\}^*$, we have $f(a,x) \neq f(b,y)$.
3. *(Concealing property.)* Let $F_N(a)$ be the distribution obtained by evaluating $f(a,x)$ for $x \in \{0,1\}^N$ chosen uniformly at random. Then we have that the ensembles $\{F_N(a) : N \in \mathbb{N}\}$ and $\{F_N(b) : N \in \mathbb{N}\}$ are polynomially quantum indistinguishable for any choice of $a, b \in \Gamma$.

When such a scheme is used, it is assumed that some *security parameter* N is chosen. When one party (the prover in the 3-Coloring protocol) wishes to commit to a value $a \in \Gamma$, a string $x \in \{0,1\}^N$ is chosen uniformly at random and the string $f(a,x)$ is sent to the other party (the verifier in the 3-Coloring protocol). To reveal the commitment, the first party simply sends the string x along with the value a to the second party, who checks the validity of the decommitment by computing $f(a,x)$ and checking equality with the committer's first message.

A quantum computationally concealing commitment scheme based on the existence of quantum one-way permutations was described by Adcock and Cleve [1]. Although the definitions in their

paper differ somewhat from ours, in particular in that they do not consider the stronger form of non-uniformity allowing an auxiliary quantum state that we require, the result can be translated to our setting. This naturally requires a stronger notion of a permutation being one-way that forbids the possibility that a quantum circuit can invert a one-way permutation using an auxiliary input.

The 3-Coloring protocol

Now we are ready to consider the zero-knowledge properties of the Goldreich-Micali-Wigderson Graph 3-Coloring protocol with respect to quantum verifiers. The protocol is as follows:

Computational Zero-Knowledge Protocol for 3-Coloring

Assume the input is a graph G with n vertices and m edges. Let $\phi : \{1, \ldots, n\} \to \{1, 2, 3\}$ be any function that constitutes a valid 3-coloring of G if one exists. Also assume a (quantum) computationally concealing $\{1, 2, 3\}$-commitment scheme is given that is described by the function f. Repeat the following steps (sequentially) m^2 times:

Prover's step 1: Choose a permutation $\pi \in S_3$ of the colors $\{1, 2, 3\}$ and strings $r_1, \ldots, r_n \in \{0, 1\}^N$ uniformly at random. Compute $s_u = f(\pi(\phi(u)), r_u)$ for each $u = 1, \ldots, n$, and send $s_1, \ldots, s_n$ to V. (Informally: commit to the coloring $\pi \circ \phi$ of G for a random $\pi \in S_3$.)

Verifier's step 1: Uniformly choose an edge $\{u, v\}$ of G and send this edge to P. (It is assumed that any dishonest verifier's message sent in this step decodes to a valid edge in G.)

Prover's step 2: Send the values $a = \pi(\phi(u))$ and $b = \pi(\phi(v))$ to V, along with the strings r_u and r_v. (Informally: reveal the committed colors for u and v.)

Verifier's step 2: Check that $f(a, r_u) = s_u$, $f(b, r_v) = s_v$, and $a \neq b$, rejecting if not. (Informally: chack the validity of the commitments and that the committed colors a and b for u and v are different.)

If the verifier has not rejected in any of the m^2 iterations, it accepts.

In the above protocol, there must be a specified choice for the security parameter N. It is sufficient to set N to be equal to the number of vertices n of the input graph for the purposes of establishing that the protocol is computational zero-knowledge.

A simulation procedure for this protocol for an arbitrary quantum or classical polynomial-time verifier V' can be constructed by simulating each iteration of the loop individually. The zero-knowledge property of the entire protocol then follows by composition.

Consider the following classical simulation for a single iteration of the protocol. The simulator uniformly chooses an edge $\{u, v\}$, and then selects a function $\mu : \{1, \ldots, n\} \to \{1, 2, 3\}$ uniformly, subject to the constraint that $\mu(u) \neq \mu(v)$. The simulator then computes commitments of the values $\mu(1), \ldots, \mu(n)$. Although μ almost certainly does not constitute a valid coloring of the graph G, the commitments of $\mu(1), \ldots, \mu(n)$ are computationally indistinguishable from commitments of $\pi(\phi(1)), \ldots, \pi(\phi(n))$ for a valid coloring ϕ when one exists. Given the commitments of $\mu(1), \ldots, \mu(n)$, along with whatever auxiliary input it may have been given, the verifier V' will choose some edge $\{u', v'\}$. In the idealized setting where one views the commitments as being *perfectly* concealing, the choice of $\{u', v'\}$ must agree with $\{u, v\}$ with probability $1/m$, independent of the actions of V'. This will not necessarily be the case when the commitments are only computationally concealing, which causes some technical complications. In case $\{u, v\} = \{u', v'\}$, the commitments of $\mu(u)$ and $\mu(v)$ are revealed, and the simulation of the current iteration is successful. As for an actual interaction, the revealed colors are uniformly distributed over the six possible distinct pairs of colors. Otherwise, the simulator "rewinds" and the entire process is repeated. By repeating the process $O(m^2)$ times, say, the simulator is very likely to obtain an iteration in which $\{u, v\} = \{u', v'\}$, representing a successful simulation.

Based on such a classical simulation, we may define a quantum simulator in a manner similar to the one in Section 4. We assume the verifier V' has a similar set of registers to before, except that now A and B store edges of G rather than just a single bit. The circuit Q is defined as before, except for the obvious changes. The unitary operator T will now represent a unitary implementation of the first part of the classical simulation just described, with the register R corresponding to all of the random bits that are needed for the simulation. This will include the random choices used for the commitments. The controlled-NOT from A to B may be taken bitwise, and the simulation is viewed as successful when all of the bits of B are set to zero (or equivalently that the classical states of A and B agree before the controlled-NOT is performed). Finally, the approximate version of the amplification lemma is applied to Q to give the final simulator for V'.

There are complications that arises in analyzing this simulator that roughly correspond to those in the classical case [11]. The two main tasks are (i) establishing that there is a negligible variation in the probability $p \approx 1/m$ that the measured output of the circuit Q indicates that a "successful" simulation has occurred, and (ii) proving that the output of Q in case of "success" is computationally indistinguishable from the output of V' when interacting with P. These tasks can be dealt with similarly, relying on the fact that the commitments are computationally concealing. A fairly straightforward *hybrid* argument establishes that a significant deviation in probability from $1/m$ in a "successful" simulation can be turned into an efficient procedure for breaking the concealing property of at least one of the commitments. Likewise, an efficient non-uniform procedure that distinguishes the admissible maps corresponding to an actual interaction and the simulator defined for a given verifier can be converted to a non-uniform procedure that violates the concealing property of the commitment scheme.

7. CONCLUSION

This paper has described a method by which some interactive proof systems can be proved to be zero-knowledge against quantum polynomial-time verifiers. A natural direction for further research is to better understand the applicability and limitations of this method. A specific question along these lines is whether the statistical zero-knowledge protocol that Goldreich, Sahai, and Vadhan [12] construct for any given honest verifier statistical zero-knowledge proof system is zero-knowledge against quantum attacks.

Another interesting topic related to this paper is the existence of quantum computationally concealing commitment schemes, which were needed for the protocol of the previous section. Such schemes follow from the existence of quantum one-way permutations [1]. The existence of quantum one-way permutations and more generally quantum one-way functions that can be efficiently computed in the forward direction by classical computers has the potential to become one of the most important questions facing theoretical cryptography if quantum computers are constructed. What are the best candidates for such functions?

Acknowledgments

I would like to thank Claude Crépeau for sharing his thoughts and insight on zero-knowledge, and for getting me interested in the problem discussed in this paper in the first place. I would also like to thank Gilles Brassard, Richard Cleve, Ivan Damgård, Simon-Pierre Desrosiers, Lance Fortnow, Dmitry Gavinsky, Dan Gottesman, Jordan Kerenidis, Hirotada Kobayashi, Ashwin Nayak, Amnon Ta-Shma, and Alain Tapp, among others, for helpful discussions on quantum zero-knowledge. Finally, I am indebted to several anonymous referees who provided very helpful suggestions for improving this paper. This research was supported by Canada's NSERC, the Canada Research Chairs program, and the Canadian Institute for Advanced Research (CIAR).

8. REFERENCES

[1] M. Adcock and R. Cleve. A quantum Goldreich-Levin theorem with cryptographic applications. In *Proceedings of the 19th International Symposium on Theoretical Aspects of Computer Science*, volume 2285 of *Lecture Notes in Computer Science*, pages 323–334. Springer-Verlag, 2002.

[2] D. Aharonov, A. Kitaev, and N. Nisan. Quantum circuits with mixed states. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, pages 20–30, 1998.

[3] G. Brassard, D. Chaum, and C. Crépeau. Minimum disclosure proofs of knowledge. *Journal of Computer and System Sciences*, 37:156–189, 1988.

[4] G. Brassard, P. Høyer, M. Mosca, and A. Tapp. Quantum amplitude amplification and estimation. In *Quantum Computation and Quantum Information: A Millennium Volume*, volume 305 of *AMS Contemporary Mathematics Series*. American Mathematical Society, 2002.

[5] H. Buhrman, R. Cleve, R. de Wolf, and C. Zalka. Bounds for small-error and zero-error quantum algorithms. In *Proceedings of the 40th Annual IEEE Symposium on Foundations of Computer Science*, pages 358–368, 1999.

[6] I. Damgård, S. Fehr, and L. Salvail. Zero-knowledge proofs and string commitments withstanding quantum attacks. In *Advances in Cryptology – CRYPTO 2004: 24th Annual International Cryptology Conference*, volume 3152 of *Lecture Notes in Computer Science*, pages 254–272. Springer-Verlag, 2004.

[7] S. Even, A. Selman, and Y. Yacobi. The complexity of promise problems with applications to public-key cryptography. *Information and Control*, 61:159–173, 1984.

[8] O. Goldreich. *Foundations of Cryptography: Volume 1 – Basic Tools*. Cambridge University Press, 2001.

[9] O. Goldreich. Zero-knowledge twenty years after its invention. Electronic Colloquium on Computational Complexity (http://www.eccc.uni-trier.de/eccc/), Report No. 63, 2002.

[10] O. Goldreich and S. Goldwasser. On the limits of nonapproximability of lattice problems. *Journal of Computer and System Sciences*, 60:540–563, 2000.

[11] O. Goldreich, S. Micali, and A. Wigderson. Proofs that yield nothing but their validity or all languages in NP have zero-knowledge proof systems. *Journal of the ACM*, 38(1):691–729, 1991.

[12] O. Goldreich, A. Sahai, and S. Vadhan. Honest verifier statistical zero-knowledge equals general statistical zero-knowledge. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, pages 23–26, 1998.

[13] O. Goldreich and S. Vadhan. Comparing entropies in statistical zero-knowledge with applications to the structure of SZK. In *Proceedings of the 14th Annual IEEE Conference on Computational Complexity*, pages 54–73, 1999.

[14] S. Goldwasser, S. Micali, and C. Rackoff. The knowledge complexity of interactive proof systems. *SIAM Journal on Computing*, 18(1):186–208, 1989. Preliminary version appeared in *Proceedings of the Eighteenth Annual ACM Symposium on Theory of Computing*, pages 291–304, 1985.

[15] J. van de Graaf. *Towards a formal definition of security for quantum protocols*. PhD thesis, Université de Montréal, 1997.

[16] L. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*, pages 212–219, 1996.

[17] J. Håstad, R. Impagliazzo, L. Levin, and M. Luby. A pseudorandom function from any one-way function. *SIAM Journal on Computing*, 28(4):1364–1396, 1999.

[18] A. Kitaev. Quantum computations: algorithms and error correction. *Russian Mathematical Surveys*, 52(6):1191–1249, 1997.

[19] A. Kitaev, A. Shen, and M. Vyalyi. *Classical and Quantum Computation*, volume 47 of *Graduate Studies in Mathematics*. American Mathematical Society, 2002.

[20] A. Kitaev and J. Watrous. Parallelization, amplification, and exponential time simulation of quantum interactive proof system. In *Proceedings of the 32nd ACM Symposium on Theory of Computing*, pages 608–617, 2000.

[21] H. Kobayashi. Non-interactive quantum perfect and statistical zero-knowledge. In *ISAAC 2003 – Proceedings of the 14th International Symposium on Algorithms and Computation*, volume 2906 of *Lecture Notes in Computer Science*, pages 178–188. Springer-Verlag, 2003.

[22] C. Marriott and J. Watrous. Quantum Arthur-Merlin games. *Computational Complexity*, 14(2):122–152, 2005.

[23] D. Mayers. Unconditionally secure quantum bit commitment is impossible. *Physical Review Letters*, 78:3414–3417, 1997.

[24] M. Naor. Bit commitment using pseudorandomness. *Journal of Cryptology*, 4(2):151–158, 1991.

[25] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[26] A. Sahai and S. Vadhan. A complete promise problem for statistical zero-knowledge. *Journal of the ACM*, 50(2):196–249, 2003.

[27] P. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26(5):1484–1509, 1997.

[28] J. Watrous. Limits on the power of quantum statistical zero-knowledge. In *Proceedings of the 43rd Annual Symposium on Foundations of Computer Science*, pages 459–468, 2002.

[29] J. Watrous. PSPACE has constant-round quantum interactive proof systems. *Theoretical Computer Science*, 292(3):575–588, 2003.